# PROJECTCANVAS Title (preliminary): Recommendations for new board games Group members: Sanni Isokirmo, Ahti Katiska, Miika Mäki Workshop # : 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| MOTIVATION 🎯***Which is the target group of our mini-project? Who is the end-user?***BoardGameGeek users / Board gamers (Depends on if we use personal user stats or not)  ***What are their objectives? What needs do we need to*** ***address with our work?***The objective is to find board game recommendations based on their interests/likes.   ***How will they benefit from this proposed solution?***They will benefit by finding new board games easily. | DATA COLLECTION 🧩***Which data sources are we planning to use?***Board Gamer Geek   ***Mention database tables, API methods, websites to scrape, etc.*** *Via the BGG API* <https://bgg-api.readthedocs.io/en/latest/modules.html> ***Which is the data management plan?***We store data as feather files in the project data directory | PREPROCESSING 🛠***What are the goals of the preprocessing pipeline?***  Have data in tidy data frame format.  ***Give some examples of data preprocessing steps. What are some possible data cleaning/wrangling methods you’re planning to use?***  -drop missing review data - only include users that have several reviews***What are some possible data transformations that could be useful?***  ***Any feature engineering necessary?***No, we have user ratings already unless we will use other features. | EXPLORATORY DATA ANALYSIS (EDA) 🔎***Look at the data!***  -we will!  ***What steps are you planning to take towards exploring and understanding better the data you have? What properties would be meaningful to summarize/visualize in this step?***-we could explore how many reviews users have left and how well they cover the boardgames (sparsity of our training matrix) -average ratings-histograms | VISUALIZATIONS 📊***List any meaningful visualizations you are planning to produce that will be useful to the end user?***  -Apart from the match score (see below), we might also display relationships in the training data for the user to explore. ***Are you planning to produce any interactive visualizations?***We will visualize the match score with a progress bar  ***If so, which types of interactivity might be useful to the end user?*** |
| LEARNING TASK 🐭 (focus on problem definition) ***Define the problem setting.***  ***Is this supervised / unsupervised / other…?***  Supervised learning  ***Classification / regression / other…?***  Regression with ranking  ***What are we planning to learn? E.g. What is the target variable / learning outcome?***  Target variables are the new board games that the user might like.  ***What variables are we using as input?***  *All revie*w scores from other users and the current user. Possibly others but we start with that. | LEARNING APPROACH 🪤 (focus on solution implementation) ***Which ML/statistical methods seem more relevant for the defined problem setting and why?***  Non-negative matrix factorization. It is a powerful recommendation algorithm.***Which evaluation metrics could be relevant?***Possibly, RMSE, Precision\_K and Recall\_K ***Is any special treatment relevant regarding how we choose to split the data or how we cross-validate?***-we will use 5-10 fold cross validation and see how it works. |  | COMMUNICATION OF RESULTS 📢 ***Which type of deliverable will benefit most the end-user? Do we choose to write a blog post, create a website, an app, or other..?***We have chosen to create a website.  ***How do we communicate best our results to the predefined target group?***-The user gets recommendations if they have any previous ratings.-Time allowing, if the user does not have an account, they can choose a genres they like, and the webpage will recommend games that they could like (this would probably be  ***Short description of your interface/workflow (if applicable).***-Web page where one can enter a username (or leave it blank)- If blank, the user can choose genres they like - Then the page will suggest 10 new games in order | DATA PRIVACY AND ETHICAL CONSIDERATIONS 🔐 (if applicable)***Are there any fairness constraints that apply to our proposed pipeline?***  -no  ***Is there a need to ask for consent during the data collection process?***  -no  ***Is there a need for data pseudonymization/anonymization?***  -no  ***Any other privacy considerations that come to mind?***-we will not store the data anywhere |
|  | ADDED VALUE🎁***Is there a possibility for added value from the data we’re planning to use?***  Yes***What is the added value?***The chance to better see new board games to fit their personal tastes. | ***How are predictions turned into added value for the end-user?***  * The added value are the recommendations |  | LEGENDWEEK 1: Data collection/preprocessing  WEEK 2: EDA & visualizations   WEEKS 3-4: Machine/deep learning  WEEK 5: Fairness & data privacy |